Time Invariant Assimilation Techniques

William Casper
(Dated: March 9, 2011)

Traditional data assimilation techniques attempt to acquire an accurate idea of the instantaneous "state" of a physical system via an interchange of observational and model probability distributions. A natural question that arises is also possible to adopt similar techniques to get a better idea of the mean or seasonal state of a system.

I. ASSIMILATING WITH CONSTANT OBSERVATIONS

A. Experimental Setup

The data assimiation routine is initialized using the POP ocean model on a 1 degree resolution ocean grid. The data assimilation is done with DART using adaptive inflation. The atmospheric forcing chosen was one of the following

NYF normal year forcing from the NOAA COREv2 data sets

AvgNYF constant forcing obtained from averaging the COREv2 monthly normal year fields

In choosing our initial ensemble, we had a lot of freedom. However, the primary possibilites can essentiall be broken down into two. Our initial ensemble was taken with 20 members and was chosen to be one of the following

NYJan an ensemble of the January 1 POP model restart files for the years 240 through 259 for a long POP spin-up run from a Levitus T, S initial condition with normal year forcing

NYEql equally spaced restart files within year 250 of a spin-up run from Levitus initial conditions with normal year forcing

The observations which we assimilate is constant in time for each experiment. The observations themselves are from tHEPlAceThATiSnOMOrE and are equally spaced over the ocean at a resolution of five degrees and at various depths. The depths of the observations we assimilate and the time interval between assimilations is varied between experiments. The specific choices of observations and assimilation intervals we used are the following

S5D05DL10 assimilation of all available observations with depths between 0 and 200 meters every five days

S5D30DL10 assimilation of all available observations with depths between 0 and 200 meters every thirty days

S5D30DL26 assimilation of all available observations with depths between 0 and 2.0 kilometers every thirty days

The particular experiments we conducted are listed in table I

TABLE I: Experimental setup for various data assimiation experiments with constant observations.

Exp #	Init	Forcing	Observations
CON01	NYJan	AvgNYF	S5D10DL01
CON02	NYJan	AvgNYF	NONE
CON03	NYEql	AvgNYF	S5D05DL01
CON04	NYEql	NYF	S5D05DL01
CON05	NYJan	AvgNYF	S5D30DL26

B. Initial Ensemble Rank Histogram

The seasonal cycle is the biggest signal. So, for the constant forcing-data DA runs (alone), it seems best to use NYEql as the IC. Most likely my bad. But, this should rectify the initial outlier nature of the observations in the constant forcing-data run.

Depending on the choice of initial ensemble, the initial rank histogram can be severely rank deficient. An explanation of this is that for an initial ensemble like NYJan, then ensemble represents a best guess to the state of the system at a specific day of the seasonal cycle. Unfortunately, the observations themselves represent the annual mean of the system. Thus the difference between the ensemble mean and the observation mean can be quite large; comparable to the size of the seasonal cycle itself. Thus in comparison to the total 28287 available observations with the S5D10DL01 observation setup, a mere 544 or 1.9 percent are actually contained by the initial rank histogram in the CON01 experiment. A plot of this histogram is included in Fig. (1) below.

On the other hand, the number of observations contained within the rank histogram has a tendancy to increase as the number of assimilation cycles increases. A plot of the percentage of observations contained within the rank histogram as a function of the assimilation time is included in Fig 3 below

C. Evolution of Ensemble Error and Spread

Riley, Please use these changed definitions for error and spread diagnostics

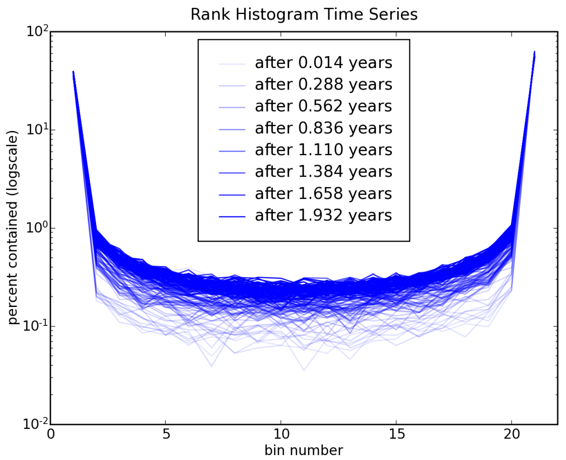


FIG. 1: A time series of rank histograms for the CON01 experiment. Because the initial ensemble is chosen with as the best guess to a specific day of the year and the observations represent the best guess to the annual mean, the ensemble mean lies well away from most of the observations. As a result, the initial ensemble is severly rank deficient with respect to the observations, containing less than 8 percent of the available observations and leading to an immediate collapse to an ensemble with less than 2 percent of observations As time increases, however, the ensemble does begin to recover.

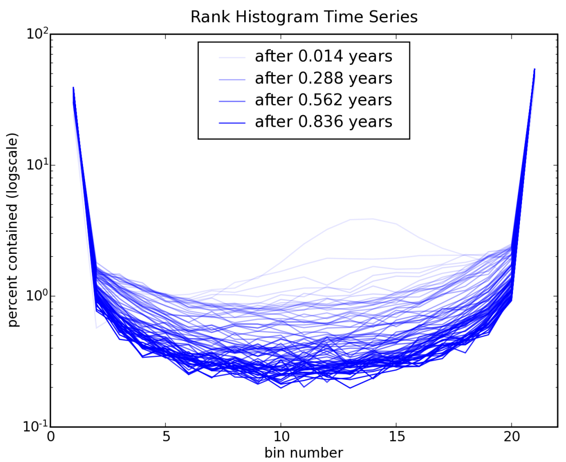


FIG. 2: A time series of rank histograms for the CON03 experiment.

$$\overline{x}_i(t) = \frac{1}{N_{ens}} \sum_{j=1}^{N_{ens}} x_{ij}(t) \tag{1}$$

$$\epsilon_i(t) = \frac{|\overline{x}_i(t) - y_i(t)|}{\sigma_i} \tag{2}$$

where σ_i is the maximum of the standard deviation of $\overline{x}_i(t)$ and standard deviation of $y_i(t)$ and which may be

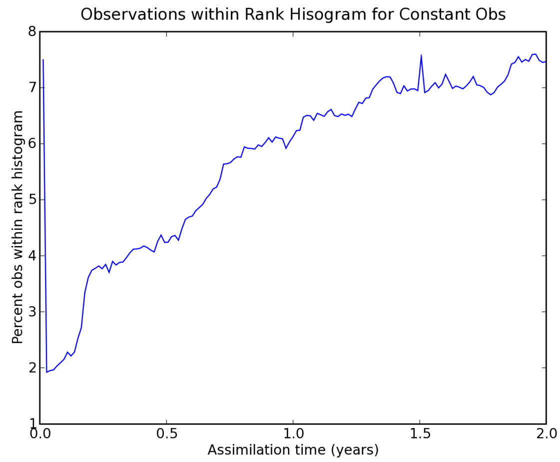


FIG. 3: The time dependence for the percentage of observations (of the 28287 available observations) contained within the rank histogram for the CON01 experiment. Here assimilation occurs once every five days with the same observations.

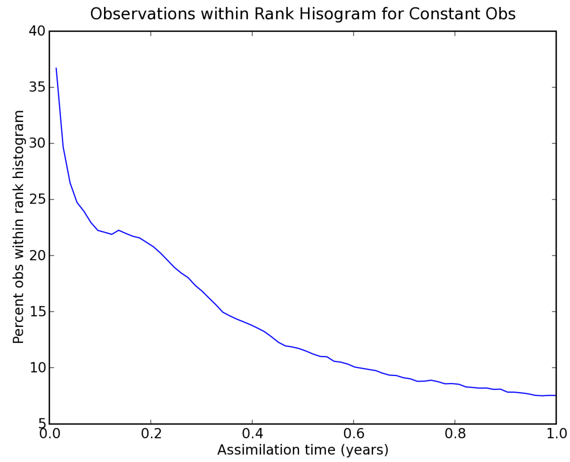


FIG. 4: The time dependence for the percentage of observations (of the 28287 available observations) contained within the rank histogram for the CON03 experiment. Here assimilation occurs once every five days with the same observations.

approximated by the standard deviation of the model estimate of that observation in a control run.

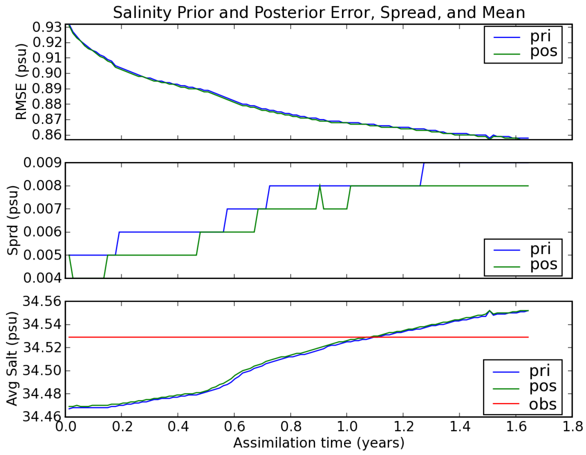
$$\epsilon(t) = \left(\frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \epsilon_i^2(t) \right)^{1/2} \tag{3}$$

$$S_i(t) = \left(\frac{1}{N_{ens}-1} \sum_{j=1}^{N_{ens}} \left(\frac{x_{ij}(t) - \overline{x}_i(t)}{\sigma_i} \right)^2 \right)^{1/2} \tag{4}$$

$$S(t) = \left(\frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} S_i^2(t) \right)^{1/2} \tag{5}$$

$$S(t) = \left(\frac{1}{N_{obs}} \sum_{i=1}^{N_{obs}} \frac{1}{N_{ens}-1} \sum_{j=1}^{N_{ens}} \left(\frac{x_{ij}(t) - \overline{x}_i(t)}{\sigma_i} \right)^2 \right)^{1/2} \tag{6}$$

A plot of these various quantites for temperature and salinity as a function of the assimilation time is included in Figure (5) below. As can be seen, the spread increases with time as the error decreases. Is this an indication that the assimlation has yet to settle down and is still making major adjustments to the ensemble state, or because an annual mean state has more uncertainty associated with it than a specific day? Most likely, the answer is the former, and the ensemble will continue to adjust toward the mean state as time increases. Another interesting observation is that the average salinity seems to overshoot the average from observations, while the temperature seems to be asymptotically approaching it. Has this to do with the smaller spread associated with the salinity, or something else?



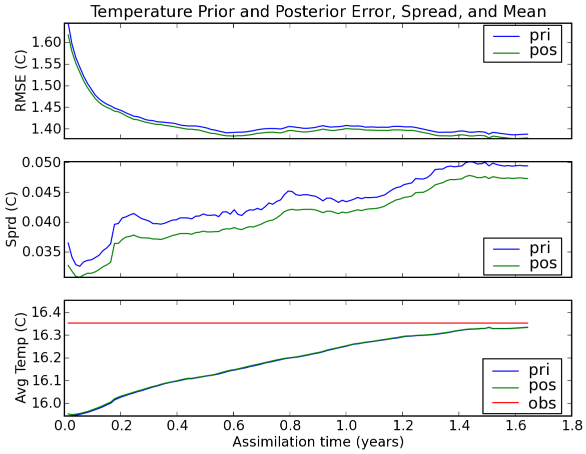


FIG. 5: The error, spread, and mean value behavior for salinity and temperature as a function of the assimilation time for the CON01 experiment. The error has a tendancy to decrease in time, while the spread increases. All quantities are calculated and averaged in the observation space.

D. Ensemble Correction and Forecast Runs